Chart Interpretation and Summarization using Intelligent Techniques

Synopsis Report of

Doctor of Philosophy

by

Ms. Prerna Ajay Mishra

(Roll No: 185000101)

Under the supervision of

Dr. Santosh Kumar

Assistant Professor (CSE)

and

Dr. Mithilesh Kumar Chaube

Assistant Professor (Mathematics)

Department of Computer Science and Engineering
International Institute of Information Technology, Naya Raipur
Raipur-493661
June 2022

डॉ.श्यामाप्रसाद मुखर्जी अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान
DR. SHYAMA PRASAD MUKHERJEE INTERNATIONAL
INSTITUTE OF INFORMATION TECHNOLOGY, NAYA RAIPUR
(A Joint Initiative of NTPC & Govt. of Chhattisgarh)

Email: iiitnr@iiitnr.ac.in
Tel: +91-771-2474040
Web: www.iiitnr.ac.in

Declaration by the Student

Date: June 15, 2022

Title of the Thesis: **Chart Interpretation and Summarization using Intelligent Techniques**

I certify that,

a. The work contained in the Thesis is original and has been done by me under the guidance of my supervisors;

b. The work has not been submitted to any other Institute for any degree or diploma;

c. I have followed the guidelines provided by the Institute in preparing the Thesis;

d. I have conformed to ethical norms and guidelines while writing the Thesis and;

e. Whenever I have used materials (data, models, figures, and text) from other sources, I have given due credit to them by citing them in the text of the Thesis, giving their details in the references, and taking permission from the copyright owners of the sources, whenever necessary.

Signature of the Student:

Name of the Student: Prerna Ajay Mishra
(185000101)

डॉ.श्यामाप्रसाद मुखर्जी अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान
DR. SHYAMA PRASAD MUKHERJEE INTERNATIONAL
INSTITUTE OF INFORMATION TECHNOLOGY, NAYA RAIPUR
(A Joint Initiative of NTPC & Govt. of Chhattisgarh)

Email: iiitnr@iiitnr.ac.in
Tel: +91-771-2474040
Web: www.iiitnr.ac.in

Declaration by the Supervisor

This is to certify that the Thesis entitled **Chart Interpretation and Summarization using Intelligent Techniques,** submitted by **Ms. Prerna Ajay Mishra** to International Institute of Information Technology, Naya Raipur, is a record of bonafide research work under my supervision and I consider it worthy of consideration for the Synopsis of the Thesis to be submitted in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy of the Institute.

Supervisor: Dr. Santosh Kumar
Co-Supervisor: Dr. Mithilesh Kumar Chaube

Date: June 15, 2022.

Abstract

Charts provide compact and coherent information visualization and are often used to obtain significant insights from complex data. Summarizing data encodings from diverse charts requires accurate interpretation of textual and graphical information from chart images. Existing information retrieval methods process and extracts pertinent information from chart images with the basic structure. However, the retrieval system fails to interpret data relationships with the complex chart layout, thus debasing the data interpretation and summaries. Creating a generalized chart data extraction system is challenging as charts pose high diversity in design and orientations. Moreover, the existing interpretation and summarization system focuses more on direct data and thus does not extract the intended message from chart images. Also, not all charts are accompanied by any contextual message, and mainly they are not addressed in the textual descriptions. People with visual impairment and learning disabilities rely on screen readers for understanding the text but are unable to comprehend charts. Thus, it is necessary to develop a chart interpretation and summarization framework that can efficiently detect, analyze, process, and interpret chart images to generate suitable concise summaries.

This work aims to overcome the gaps in chart processing and interpretation by performing a substantial literature survey across the board, evaluating various approaches over different chart types, and conducting a profound pilot study to understand the needs of the visually impaired user. In this vein, four distinct problems are formulated under each module of chart processing targeting a broader coverage of document layout analysis and chart detection, similar-structured chart classification, data interpretation, and summary generation. This work mitigates these problems by using six style-independent supervised approaches for end-to-end chart summary generation. A cross-domain supervised approach detects charts from various document layouts. However, the detection model still overlooked a few types of charts due to its similar structural design and patterns. This problem is surmounted by fusing chart microstructural features based on chroma effects with the low-level feature representation. This reduces the within-class and between-class classification problem of chart images. This method is further enhanced by fusing heterogeneous dissimilar and similar structural features with the learning model to classify charts irrespective of their chroma effects, background patterns, and appearances. Another challenge is to extract variable-sized chart components irrespective of the location and structural layouts. A semi-automated supervised approach is developed that extracts direct and derived data from chart images irrespective of the style and structure, builds global relationships between chart components, and analyzes the underlying data for complete chart interpretation. Subsequently, a context-aware template-based summarization approach generates short summaries uncovering the intended message from the chart images by utilizing all the extracted and interpreted data. This method is further enhanced to obtain automated summaries from charts by using an end-to-end chart information system that generates detailed and short summaries in English and Hindi languages.

Experimental results and comparison with state-of-the-art benchmark methods are performed by employing the methods over various chart datasets comprising different types of charts. Results show that the proposed approaches achieve a performance rate of 95%-99%, surpassing the existing methods by 5%-10%. The summary generation system generates concise and semantically correct summaries uncovering intended meaning from chart images. The pilot study and task-based evaluation show that the generated summaries are short yet detailed enough to cover the intended message of the charts and are easily comprehended.

Keywords: *Chart Interpretation, Chart Summarization, Human-Computer Interaction, Assistive Technology, Intelligent Techniques.*

Chapter 1: Introduction

For a decade, every document, such as scientific articles, books, reports, magazines, and newspapers, comprises statistical charts as an effective medium of data presentation and visualization [1]. On unified with descriptive text, charts engage the audience with the perceptually efficient representation of data, while descriptive text proffers guidance to the readers by explaining additional contexts of the data. Summarization is generally stated as the gist of message content or subject matter. Machine-vision techniques allow us to extract meaningful information from the chart detected objects to understand their relation for complete summarization results. For representing the extracted information from chart images, the encoding and decoding of information are needed. Machine-based interpretation and summarization are still tedious tasks for processing documents. Thus, extracting hidden meaning from chart images becomes necessary to apprehend the global message on underlying data depicted by charts [2]. Recent works in automatic chart reading and interpretation opened a new passage for researchers [3].

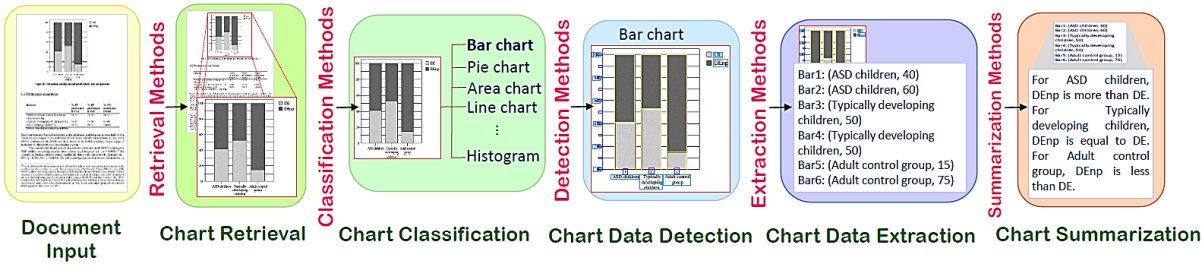


Figure 1: Workflow for chart processing system

1.1. Chart Processing System

The chart processing pipeline is organized around a few basic steps (shown in Fig. 1),

1. *Chart Detection and Classification*: Chart detection is the fundamental step of the chart processing pipeline. Chart detection finds and localizes the charts present within the input document images. Two strategies are followed for chart detection. First, detecting figure blocks from documents and then classifying the figure blocks as a chart or not a chart [4]. Second, often called multi-class chart detection detects the presence of charts from documents and classifies them as per their types.
2. *Chart Data Extraction and Interpretation*: The interpretation of charts depends on the data extraction from chart images. Data extraction is performed in a few steps, first detecting and extracting textual components, detecting graphical components with shapes, size, and colors, and identifying the relationship between components [5].
3. *Chart Summary Generation*: Summaries are the concise high-level description of the chart images covering the intended message of the image [6]. Correct summaries require appropriate details of textual and graphical components. Summaries cover available textual details, extracted details, and intended messages represented by graphical segments [7].
4. *Chart Visualizations:* Even a classic elementary chart requires a little revamp to increase visualization refinement and reduce noise and design bias. The designing approaches further allow human intervention to customize the chart as per their favorable choices, export the data and design parameters for analysis, and alter the design externally [8].

1.2. Motivation and Objectives

Even though state-of-the-art methods are powerful in specific domains such as classification, segmentation, detection, retrieval, and indexing tasks, producing high-level understanding based on the intended underlying data and summarizing it in semantically correct ways is still challenging. Moreover, giving a system the ability to summarize the chart images is still not robust enough for the chart structure and types diversity. Also, many resources are provided in a format that is not readily accessible to everyone. Individuals with sight impairments have limited access to multimodal

documents. The document's text can be quickly delivered via screen reader programs, but the document's information graphics pose serious problems. Accessibility guidelines propose that alt text (text that conveys the content of a graphic) be included with information graphics.

The primary objective of the work is to generate short summaries from charts that portrays the intended message of the chart image. These messages or summaries can be conveyed to the user using their spoken language. The second objective is to generate short and straightforward summaries of scientific charts from various types of charts, irrespective of the structural variations of the charts. The third objective of the work is to provide an assistive learning facility to people with visual and learning impairments using which they can independently perform their daily learning activities in the same as sighted users.

1.3. Applications

Chart interpretation and summarization have various applications despite being at the nascent stage. A few of them are stated below.

- Visual Question Answering: The visual question answering (VQA) system provides an appropriate response/answer from a chart for the input questions [9].
- Chart Accessibility: Accessibility of chart images using speech synthesis provides assistive aid for chart learning [10].
- Chart Captioning: Chart captions portray data trends and uncover latent abstract meaning from the image.
- Chart Indexing: Summaries, extracted chart data, chart type, abstract meaning, properties, and role classification of data are also important for automatic image indexing, which can be applied in Content-Based Image Retrieval (CBIR).
- Chart Re-designing: Charts could be redesigned or revamped to ease understanding, remove design biases, simplify data presentation, and improve the visual aesthetics [11].

1.4. Contribution to the Thesis

Based on the motivation and objectives, research contributions of the proposed work are listed below:

1. A cross-domain refined features-based document object detection approach is proposed, which accurately represents and aligns both the document's spatial and discriminative contextual features and reduces the classification misalignment between source and target domain.
2. A low-level heterogeneous feature learning approach using chroma effects within the chart microstructural features is proposed to reduce within-class and between-class classification problems within chart images.
3. A representation-based learning approach is proposed that forces the classification model to learn both the dissimilar and similar features, making the model more prone to discriminatory graphical features of similar structured chart images.
4. A semi-automated data extraction and interpretation pipeline is proposed to extract direct and derived data from chart images, decodes the inter and intra-component relationships, and interprets the underlying data from chart images irrespective of their styles and appearances.
5. A partial template-based summary generation approach is proposed for uncovering the intended chart message by articulating both direct and derived details from the chart image. Summaries are generated in the English language.
6. A style-invariant end-to-end automated system is proposed that learns graphical-textual relationships based on the semantic embeddings of the chart components and produces short summaries using relational reasoning between chart elements.
7. Three chart datasets comprising various chart types of diverse appearances and style is proposed.

Chapter 2: Related Literature

A brief outlook of the existing literature is tabulated in Table 1, depicting the broader domain, aim and objective achieved, models/methods used to achieve the proposed aim, and comments describing the limitations and novelty of the work.

Table 1: Brief overview of the literature reviewed based on each domain

Domain	Aim	Approach	Focuses on	Comment
Chart Detection	Detecting scientific images from documents	CNN [12]	Textual regions	Focused only on textual elements. Misclassifies captions and table data. Avg. accuracy is 80-85%.
		Geometric-based [13]	Graphical regions	Assumes certain locations of images. Overlooks images centered in the document.
		Region proposal Networks [14]	Textual and graphical regions	Analyses complete document layout. Misclassifies similar-looking graphical elements.
		Fusion-based [15]	Graphical regions and their captions	Overcomes alignment issues. Overlooks graphical elements between paragraphs.
Chart Classification	To classify chart images as per their types.	CNN [16]	Different structured chart types	Misclassifies similar structured charts.
		Low-level features [11, 17]	Chart types of basic layout	Suitable only for simple structured charts.
		Heuristic features based [18]	Simple structured chart types	Fails for charts having variations.
Chart Interpretation	To extract textual data from chart images. Methods only for bar chart, line chart and scatter chart are available in the literature.	Deep Learning [19]	Bar, Line, Pie	Extracts all data from images. Extracts relevant data for graphical segments.
		Heuristic-based [12]	Area, Bar, Line, Pie, Scatter	Assumes a certain orientation of graphical segments. Does not extract data for graphical segments. Fails to relate legend and graphical segments.
		Fusion-based [20]	Bar, Line, Pie, Area,	Fusion-based methods overcome alignment issues and need separate methods for each chart type.
Chart Summarization	Generating summaries from bar and line chart.	Template-based [21]	Bar and Line	Follows a template to produce a summary. Does not generate intended meaning. Focuses more on data present in the image.
		Automated [22]	Bar, Line, Pie, Area	Interpret the intended message of charts. Generates short summaries. Robust towards various chart types.

2.1. Limitations of Existing Work

As per the literature reviewed, a few issues were identified, such as,

1. Literature does not discuss interpretation in the absence of textual segments in chart images and what alternative methods could be incorporated [2, 3].
2. No work in literature generates a summary from derived data of charts. The work focuses only on direct data available on the chart.

3. The same methodologies cannot be employed to extract and recognize all types of charts because charts have varied physical appearances and characteristics [11].
4. No work has been suggested that recognizes different chart types and sub-types having similar features and appearances, such as horizontal and horizontal bar chart-grouped bar charts and vertical bar chart-histogram.
5. The summary generation system is proposed only for the line charts and bar charts. Subtypes such as grouped bar charts and horizontal bar charts are not considered.
6. Summary generation systems describe only the outlook of the charts. Systems also generated summaries either in a table or XML format. All of them lacked interpretation of an intended message.

2.2. Available Chart Datasets

Few chart image datasets are publicly available for document layout analysis or object detection, classification, and summarization tasks shown in Table 2. However, very few datasets cover different chart types and subtypes, while some focus on specific objects.

Table 2: Available public chart datasets

Datasets	Size	Details
DocBank [24]	5,00,00 images	One of the largest document datasets of 12 categories of annotations.
PubLayNet [25]	3,64,262 images	Majorly used for document layout analysis having 13 categories of annotations.
FiqureQA [9]	100,000 images	A well-known visual reasoning dataset has four chart types.
ReVision Corpus	5,424 images	Corpus used by Savva et al. [11] comprises 10 different chart types.
CHART Infographics [26]	2,02,550 images	An ICDAR2019 Competition on Harvesting Raw Tables from Infographics having 6 classes of charts.

Summary: Literature does not contribute much towards chart summarization, even though some works were focused on generating data from chart images. In contrast, others were focused on generating short summaries in sentences, tables, or XML format, through similar framework architectures, with small variances, such as convolutional neural network + recurrent neural network. There is also a strong need to address the challenging issues of chart processing by stepping away from heuristics, hard assumptions, and constraints while progressing towards robust, style-invariant, and scalable methods.

Chapter 3: Chart Detection and Document Layout Analysis

3.1. Background

Document layout analysis and object detection is the initial step of chart summarization. It includes decomposing a document page or document layout into its structural units (e.g., figures, tables, paragraphs). However, variations of style, scale, and appearance of a different object in the documents in different domains severely affect the performance of detection models. Moreover, the highly varied and diverse document structure used for different applications also causes a severe challenge for the learning model to discriminate styles and structure. In addition, objects in the documents are not fixed to a number; thus, even minute differences in the structure make it difficult for any document recognition and analysis model to recognize various identities, thereby influencing the model's performance.

3.2. Proposed Method: Cross-domain Approach for Layout Analysis and Object Detection[1]

A cross-domain refined features-based document object detection model is proposed, which is amalgamated with feature attention mechanism i) Document Object Attention (DA) module, which is a refined-feature extraction and attention mechanism, and ii) Classifier Alignment (CA) module, which reduces the classification misalignment between source and target domain. DA module is intermixed with deformable backbone architecture to generate semantic-aware contextual features. The deformable detection model detects and classifies object proposals obtained from the RPN. CA module is responsible for reducing the false or misclassified prediction performed by the detection model while obtaining the correct classification for each object class across the source and target domains. The classification here can also be understood as role classification, i.e., predicting each object's accurate label (role) in a document. For example: predicting the textual objects like paragraph, list, formula, header, title, footer, page number, and others, and non-textual objects such as tables, charts, and diagrams. Fig. 2 shows the proposed cross-domain model. The ResNet-50 network is adopted as the backbone network. The model is trained on company annual reports and tested on standard documents, magazines, and scientific reports of various layouts.

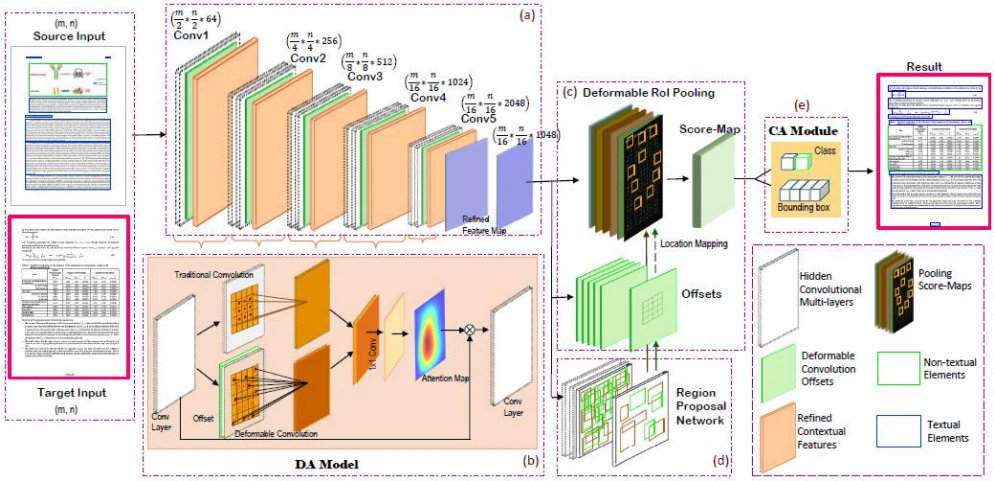


Figure 2: Overview of the proposed cross-domain detection model

3.3. Experimental Analysis

To perform cross-domain adaptation for object detection, source and target datasets of different domains were considered, further represented as Source → Target.

3.3.1. Datasets Used

- DocBank [24]

[1]P Mishra, S Kumar, M K Chaube, "Adaptive Learning for Cross-domain Document Illustrative Detection Using Dual Attention Mechanism", Springer Journal of Multimedia Information Retrieval.

- IIITAR-13K [27]
- ICDAR POD 2017 [28]
- MixedBag
- DSSE [33]

3.3.2. Hyper-parameter Settings

The experiments were performed on a workstation with 64GB RAM, Intel Xeon - Deca core processor, and dual GeForce GTX 1080TI GPU. The hyper-parameter setting used to evaluate the model over different datasets is, network optimizer = stochastic gradient descent (SGD), learning rate=0.0004, momentum = 0.9, IoU threshold τ = 0.6, iterations = 50,000, weight decay = 0.0005. Training data is divided into source training samples with labels and bounding boxes and target training samples without labels and bounding boxes.

3.3.3. Results Obtained

Table 3 shows the performance rate in terms of F1 values obtained by the proposed model with the CA module (Faster RCNN+DA+CA) and without the CA module (Faster RCNN+DA) over different datasets. Accurateness of class 'charts' was determined if the model could predict the actual chart type instead of just the main category. The performance results of a few source → target series were obtained on document datasets in comparison with benchmark models.

Table 3: Prediction performance (%) of the proposed model for each document category over different datasets (I. With CA Module, II. Without CA Module)

Labels	DocBank →MixedBag		DocBank→ICDAR		DocBank→ DSSE		IIITAR→MixedBag		IIITAR→ICDAR		IIITAR→DSSE	
	I	II	I	II	I	II	I	II	I	II	I	II
Title	95.2	88.8	97	89.8	95.1	90.4	97.5	89.7	96.7	86.3	98	92.3
Author	95	88.5	97.2	87.7	94.5	85.6	96.5	91.2	95	88.7	96	90
Abstract	94.6	79.2	96	88.7	94.2	86.3	93.5	82.3	96.5	87.9	94.7	89.9
Keyword	-	-	96.1	91.4	92.4	87.1	-	-	96.5	90.8	92.8	91.2
Copyright	-	-	-	-	95.8	84.3	-	-	-	-	92.4	86.3
Date	96	94.3	96.5	88.2	96	91	94.5	86.2	96	86.3	94.5	89.6
Header	96.5	84.9	95	91	96.2	91.2	95.5	89.6	96.6	92.4	93.4	84.7
Footer	96.4	84.7	95.5	89.6	95.1	92.5	96.7	91.4	95	91.9	92	86
Page number	96.7	95	96.4	90.9	95.9	90	95.5	92.3	94.2	92	93.9	91
Paragraph	95.9	94	96.3	90.8	97.1	87.9	96	92.9	96.5	92.6	93.4	86.7
List	95.3	68.7	93.5	72.4	95.6	81.3	96	90.5	93	87.6	91.5	82.4
Glossary	94.9	86.4	-	-	92	84.3	95	89.9	-	-	92.4	85
References	96.7	91.7	96.5	85.5	95.6	89.4	96.3	91.2	95.7	88	93.1	88
Caption	97.9	90.3	94.6	89.9	97.1	91.2	95.4	90.7	94.9	86.5	93.9	91.7
Equation	97.8	89.8	95.2	90.2	94.8	86.2	95.9	91.4	96.6	86.2	94.4	90.7
Section Title	96.2	88.4	96.7	90.3	97.1	90.9	96.6	90.7	94.2	91.5	96.4	92.2
Declaration	93.6	86.9	-	-	-	-	96.3	86.7	-	-	93.4	91.1
Acknowledgement	94	68	-	-	92.7	86.4	93	85.7	-	-	92	82
Biography	96	90	-	-	91.9	85.2	95.5	89.2	-	-	94	91
Chart	97	87.6	96.2	91.5	95.5	90.1	96.8	90.8	98.1	91.4	95	85.6
Table	97.1	91.1	94.6	90.7	97	92.7	96	91.3	96.7	89.4	96.4	92.6
Figure	97.9	86.6	96.6	91.3	95	87.3	95.1	92.7	95.4	87.6	95.9	94.8
Author Image	98.5	90.9	-	-	94.4	88.2	94.7	85.6	-	-	96	87.2

- Not available in the dataset

Despite MixedBag being the synthetic dataset, the proposed model has achieved good detection results with all other datasets. Few cases were seen where sub-section titles were considered as a paragraph. This was because paragraph contents immediately followed the sub-titles (a sample is shown in Figure 3-a). Detection is categorized into three types, 1) Correct in which prediction overlaps with ground truth is greater than 0.5, 2) Mislocalize in which prediction overlaps with ground truth is between 0.3 and 0.5, 3) Background in which prediction overlaps with ground truth is less than 0.3. Top P predictions were considered for each element category, with P being the total number of ground truths for a category. The results shown in Figure 4 shows that CA and DA modules have increased the correct detections while reducing the false positives compared to Kai et al. [2] from 45% to 52%, which is an increase of 7%.

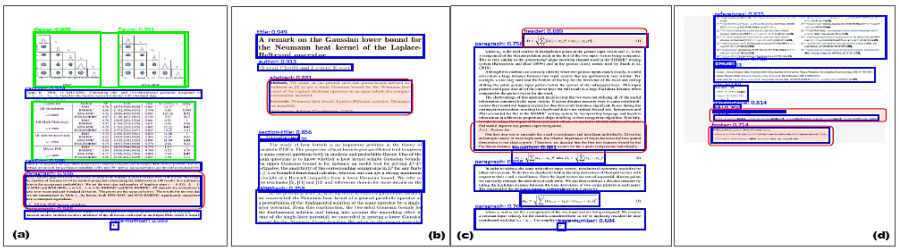


Figure 3: Error detection and misclassification (errors highlighted in red, non-textual objects highlighted in green, textual objects highlighted in blue).

Table 4: Performance comparison (AP/oLRP) of textual and non-textual objects from Source→Target datasets

Dataset→	DocBank→MixedBag		DocBank→ICDAR		DocBank→DSSE		DocBank→IIITAR	
Models↓	TO	NO	TO	NO	TO	NO	TO	NO
FRCNN	82.5/60.8	83.8/59.8	83.1/56.8	81.6/58.7	80.9/61.4	84.6/55.9	85.8/57.4	86.2/55.4
FRCNN+DA	*87.6/50.7*	*89.2/48.1*	*88.3/50.8*	*89.1/49.8*	*87.9/50.3*	*89.7/48.3*	*90.3/49.6*	*90.1/50.4*
FRCNN+CA	*88.8/55.2*	*89.9/46.3*	*87.4/48.5*	*87.8/52.3*	*88.1/54.9*	*88.0/50.6*	*88.4/53.3*	*89.1/48.7*
FRCNN+DA+CA	*94.1/45.2*	*93.4/42.6*	*93.9/40.4*	*94.9/39.8*	*94.3/41.3*	*93.4/42.6*	*93.8/41.8*	*94.0/41.3*
Kai et al. [2]	84.4/51.7	87.4/51.6	85.6/54.9	84.8/52.4	82.8/59.6	87.4/54.3	86.8/57.2	86.4/56.8
Finn et al. [29]	85.3/52.2	86.2/55.4	83.1/53.2	83.6/50.7	88.4/59.6	85.9/55.3	84.4/56.7	89.8/58.2
Kai et al. [30]	84.5/53.4	83.8/54.7	84.5/55.4	82.1/52.5	84.3/56.2	85.4/55.1	82.8/58.9	85.2/56.4
Singh at al. [13]	84.5/52.6	85.7/53.7	84.5/52.6	86.2/53.7	86.3/51.7	88.7/52.3	89.3/50.1	88.7/51.6
Dataset→	IIITAR→ICDAR		IIITAR→MixedBag		IIITAR→DocBank		IIITAR→DSSE	
Models↓	TO	NO	TO	NO	TO	NO	TO	NO
FRCNN	84.1/56.6	87.8/55.0	85.6/55.7	83.2/59.7	81.4/57.8	80.3/60.3	82.6/64.2	88.4/59.1
FRCNN+DA	*88.8/49.7*	*90.4/48.3*	*90.8/48.4*	*89.6/50.4*	*90.7/49.9*	*88.2/51.8*	*84.4/60.7*	*90.4/55.4*
FRCNN+CA	*91.3/48.4*	*88.3/45.2*	*87.6/45.7*	*85.8/54.9*	*88.3/50.7*	*85.0/55.5*	*86.2/55.1*	*91.7/52.3*
FRCNN+DA+CA	*94.6/41.4*	*94.9/40.8*	*92.3/43.1*	*93.4/42.2*	*94.4/40.1*	*94.9/41.6*	*92.6/48.8*	*96.0/43.3*
Kai et al. [2]	84.5/51.0	89.9/49.1	87.2/52.2	86.4/53.8	85.4/53.6	84.7/57.9	83.4/55.8	85.7/50.8
Finn et al. [29]	78.4/60.8	82.6/59.8	80.6/58.6	79.8/60.1	78.2/59.7	75.9/61.6	82.6/51.2	91.3/49.7
Kai et al. [30]	85.4/52.6	86.1/51.7	88.5/51.8	87.4/52.3	82.1/56.9	81.6/58.8	86.3/54.5	83.4/55.6
Singh at al. [13]	86.9/50.8	89.6/49.8	89.4/49.6	88.8/51.7	89.7/48.6	87.1/53.3	82.3/54.0	87.8/54.0

Abbreviations: TO-Textual objects, NO-Non-textual objects

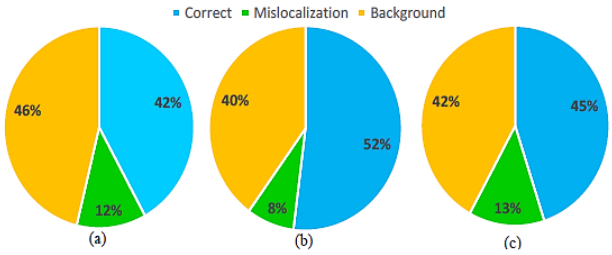


Figure 4: Error Analysis of Highest Confidence Detections. The mean percent value across each category is illustrated. (a) XDOD+CA, (b) XDOD+CA+DA, (c) Kai et al. [2]

Summary: To overcome the issue of data distribution in document datasets, a cross-domain document object detection approach fused with Document Object Attention (DA) and Instance Level Classifier Alignment (CA) modules are discussed. Experimental results and their analysis portray the improved efficiency and effectiveness of the proposed approach, giving a high detection rate of more than 94% over various document datasets.

Since the model correctly role classified almost all document objects, there were a few instances where chart components were misclassified where sub-types were classified as parent types due to their similar structural properties. In this context, classification approaches for learning the structural diversity of charts have been proposed in the Chapter 4.

Chapter 4: Chart Classification Based on Heterogeneous Features

4.1. Background

It is challenging to acquire a good feature representation of an image because each chart image has its heterogeneous features, which makes it different from other images. Heterogeneous differences are the smallest features concerned with the precise, unique, and distinct details. These heterogeneous features in the chart image make it difficult for a classifier to recognize identities, leading to the miss-classification of detected charts. As per the literature reviewed, existing methods do not classify charts considering their covariates and variations, i.e., it fails to classify within and between class problem. Chart covariates are continuous variables that are observed but cannot be manipulated, thus, influencing the outcome of the results. Covariates such as the orientation of charts, structural appearances, change of data, and varied chart style are primary parameters influencing the performance of the system [16].

Recognizing charts even on homogeneous features and variations becomes difficult [16]. These parameterized feature variations cause difficulty in classifying charts, thereby giving a poor performance because current classification methods are not adequately utilized to transform discriminatory features of charts. Moreover, existing feature extraction methods focus only on homogeneous features leaving all minute features for chart images. This causes a loss of data, which can be crucial for chart interpretation and analysis.

4.2. Proposed Method: Classification based on Low-level Heterogeneous Microstructural Features[2]

It is assumed that many common and uncommon microstructural features form every image. These structures are part of the whole multi-channel image. Extracting these microstructures and defining them effectively can be used as a core mechanism to compare and analyze various images. This method is called as heterogeneity index. The method is a four-stage process (as shown in Fig. 5). First, the chroma effects and illumination is detected. Second, features pertaining to these effects are extracted. Third, all the separate microstructures are integrated to form index features. And lastly, these microstructures are fused with local feature detector and descriptor methods. HI can define various combinations and distributions of structures; Thus, it can distinguish between colors, textures, structural layout, and illumination details.

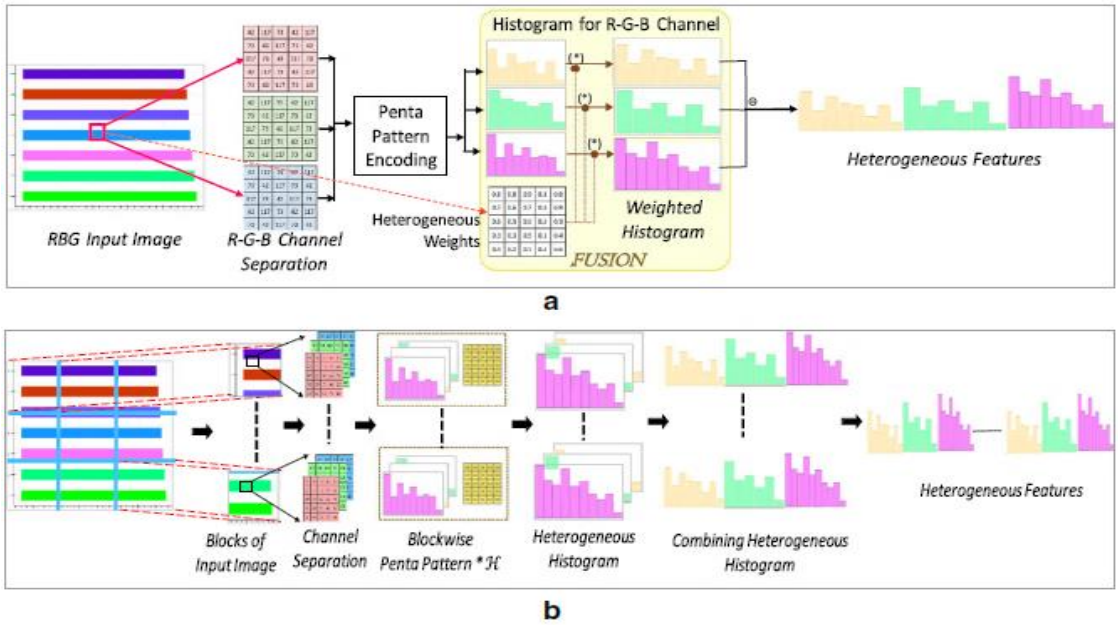


Figure 5: Flow diagrams for a) Feature extraction process of HLPP, b) Feature extraction process of MB-HLPP

[2]P Mishra, S Kumar, M K Chaube, "ChartFuse: a novel fusion method for chart classification using heterogeneous microstructures", Springer Multimedia Tools and Applications, 80, no. 7, 2021. https://doi.org/10.1007/s11042-020-10186-z.

4.3. Experimental Analysis

This section discusses the results and analysis of the performance rate of the proposed method in comparison with existing state-of-the-art feature extraction methods on different chart datasets.

4.3.1. Datasets Used

- FigureQA [9]
- Web-collected Chart Dataset
- Handcrafted Chart Dataset: For better evaluation of diverse chart types, a handcrafted chart image dataset is developed with ten different chart categories (same as the web collected dataset given above). Each category has 100 images per category, a total of 10,000 images.

80% of the dataset was used for training with 5-fold cross-validation, while 20% of the database was used for testing. A support vector machine (SVM) was employed for the classification process.

4.3.2. Results Obtained

Table 5 depicts the performance rate of the method for datasets used for evaluation. Results obtained for FigureQA and the handcrafted dataset are good. Results for web collected images are comparatively less because web collected images were of varied resolutions. Table 6 depicts the performance of the proposed method for different datasets. Inspite of variance in colors, appearances, and characteristics, the proposed method is robust to all types of charts giving the recognition accuracy of 95.96% and 97.68%, respectively.

Table 5: Performance rate (%) of the proposed approach for different database

Method	Database	Precision	Recall	F Score	Accuracy
HLPP	Web collected	90.63	90.72	91.04	90.35
	FigureQA	92.21	92.90	92.90	92.9
	Our dataset	95.96	95.93	95.94	95.91
MB-HLPP	Web collected	91.47	91.18	90.95	91.03
	FigureQA	94.83	94.80	94.79	94.80
	Our dataset	97.68	97.66	97.67	97.60

Table 6: Performance rate (%) of individual categories for different datasets

Datasets	***Web-collected Dataset***				***Handcrafted Datasets***				***FigureQA***			
Parameters	**Precision**		**Recall**		**Precision**		**Recall**		**Precision**		**Recall**	
Chart Types	HLPP	MHLPP	HLPP	MHLPP	HLPP	MHLPP	HLPP	MHLPP	HLPP	MHLPP	HLPP	MHLPP
Horizontal Bar	86.87	88.59	89.45	90.85	92.2	94.8	94.66	97.33	92.82	96.85	90.5	92.5
Vertical Bar	87.43	92.42	88.79	90.18	92.76	98.63	94	96.66	92.68	95.12	95	97.5
Line Chart	91.98	93.79	91.46	92.18	97.31	100	96.67	98.66	96.44	95.07	95	96.5
Pie Chart	92.64	93.11	91.46	91.52	97.97	99.32	96.67	98	90.45	94.87	90	92.5
Scatter Chart	89.33	89.81	89.45	90.18	94.66	96.02	94.66	96.66	92.15	92.23	94	95
Histogram	93.3	91.12	90.79	90.85	98.63	97.33	96	97.33	-	-	-	-
Donut Chart	92.64	92.44	91.45	91.52	97.97	98.65	96.66	98	-	-	-	-
Area Chart	91.98	93.11	91.45	92.18	97.31	99.32	96.66	98.66	-	-	-	-
Stack Area Chart	90.67	89.84	90.79	90.85	96	96.05	96	97.33	-	-	-	-
Bubble Chart	89.47	90.5	92.12	91.52	94.8	96.71	97.33	98	-	-	-	-
Average	**90.63**	**91.47**	**90.72**	**91.18**	**95.96**	**97.68**	**95.93**	**97.66**	**92.91**	**94.83**	**92.90**	**94.80**
Abbreviations: HLPP:- Heterogeneous LPP, MHLPP:- Multi-block Heterogeneous LPP												

Summary: The heterogeneity index was built based on the similarity of chroma effects and its intensity in the microstructures. The LPP histograms are extended by fusing HI with Penta patterns, namely HLPP and MBHLPP. It was also noticed that feature extraction methods were more receptive to minute heterogeneous features. Results showed that this method gave better results even with highly variable, diverse and heterogeneous chart images.

Chapter 5: Chart Data Extraction and Interpretation

5.1. Background

A method to recover encoded information from chart images is termed reverse-engineering of charts. Information in charts is of two main types: textual and graphical information. Thus, data extraction from charts indicates both textual and graphical information extraction. Textual information adds details to the visualization to provide pertinent detail and includes several textual information classes such as chart title, legend, axis title, axis label, and data label. This textual information varies in its character strings (numeric or alphabets) and locations. Graphical information presents the notion of perceptibility of the data in a graphical manner using vivid shapes and patterns and is the main component of charts. Extracting information from charts requires three essential steps: first, identify each textual and graphical component in the chart image; second, extract the available information from the charts; and third, analyze and interpret the extracted data.

5.2. Proposed Method: Semi-automated Pipeline for Data Extraction and Interpretation[3,4]

The data extraction pipeline is divided into three parts, separating graphical components, detecting and extracting direct data, processing graphical components for derived data extraction, and interpreting the relation between the components.

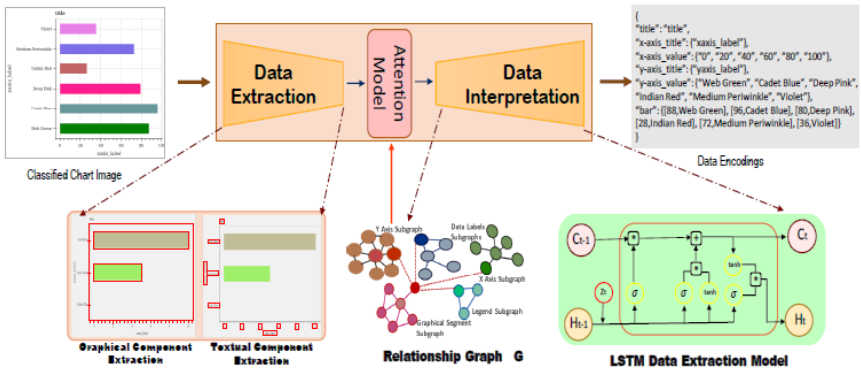


Figure 6: Semi-automated system for chart data extraction and interpretation

5.2.1. Textual Component Detection and Extraction

Early features and later features from the CNN map are fused. The region proposal network (RPN) is provided with feature maps of different stages instead of feature maps after conv5. After conv2, conv3, conv4, and conv5, the feature maps were combined through a 3 × 3 convolution with stride=2. After convolutions, the feature maps are summed up and fed into concatenate operation to form the final fused feature maps, as shown in Fig. 7. The proposed text localization model is based on the Faster R-CNN using the ResNet50 model as a learning architecture.

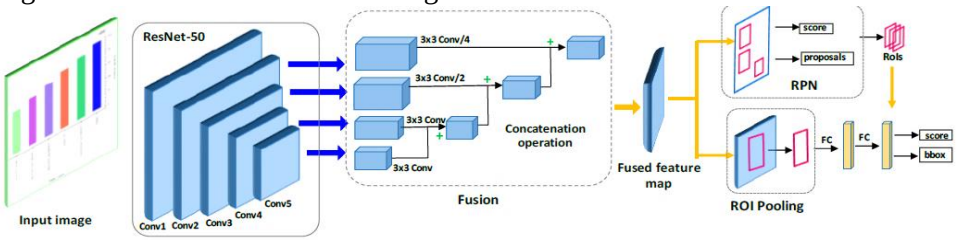


Figure 7: Text detection network

5.2.2. Graphical Component Detection

Different methods are used to detect graphical components from various types of charts. Each component is mapped with its corresponding legends, as depicted in Table 6.

---

[3]P Mishra, S Kumar, M K Chaube, "Evaginating scientific charts: Recovering direct and derived information encodings from chart images", Springer Journal of Visualization, 2021. https://doi.org/10.1007/s12650-021-00800-z.

[4]P Mishra, S Kumar, M K Chaube, U Shrawankar, "ChartVi: Charts Summarizer for Visually Impaired", Elsevier Journal of Computer Languages, Volume 69, April 2022. https://doi.org/10.1016/j.cola.2022.101107.

Table 6: Methods used for graphical component detection for different chart types

Chart Types	Legend Mapping	Graphical Segments
Pie Chart	Color of the legend box is embedded as a color-tag with each legend. Color-tag of both legend box and segment is matched, and mapped.	• Hough circle detection for segments • Transition pixels are regarded as a border point
Area Chart, Stacked Area Chart		• Canny operator is used to detect boundaries. • Histogram based color segmentation approach was used to detect areas.
Line Chart		• Canny operator fused with microstructures is used to detect boundaries. • Pearson's correlation + 8-way connectivity is applied to detect line patterns.
Histogram, Bar Chart		• Hough line transform + 8-way connectivity for boundary detection. • H channel (from HSV color space) histogram is used to obtain bars, followed by binary conversion (setting black for colors, white for others).

5.2.3. Data Interpretation

The relationship between textual-textual and textual graphical components decodes the hidden details of the chart image. These relations are represented as a graph. It comprises each component as a node and the relation between elements as edges. This forms global relations and local relations (i.e., which is obtained by decomposing the global graph into subgraphs focusing on individual segments such as y-axis-, x-axis, legend) between components as shown in Fig. 8.

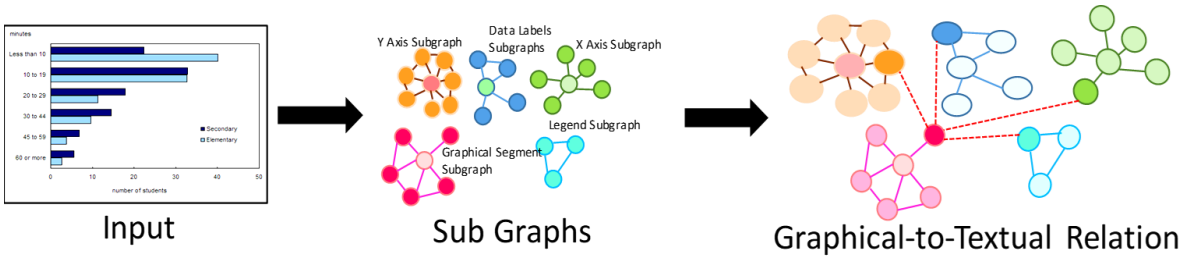


Figure 8: Example of graphical-to-textual component relation of the input bar chart

5.3. Experimental Results and Analysis

The influence of the relationship graph on qualitative and quantitative results is analyzed.

5.3.1. Datasets Used

- FigureQA [9]
- Web-based Charts

5.3.2. Hyper-parameter Settings

The hyper-parameters settings used to obtain shown results is, learning rate = 0.0001, epochs= 800, optimizer=Adam, loss function = MSE. The experiments were performed on a workstation with 64GB RAM, Intel Xeon - Deca core processor, and dual GeForce GTX 1080TI GPU.

5.3.3. Results Obtained

Fig. 9 shows the component detection and role classification of various types of charts. Table 6 shows the accurate rate of the extraction result for both textual and graphical encoding. It can be seen that, with the influence of relationship building, the model performs better for charts with varied styles and patterns, extracting rich encodings. Table 7 shows the accuracy rate for these parts measured under four parameters* such as (1) boundary detection, (2) axis detection, (3) chart-specific component, and (4) coordinate detection, in comparison to benchmark methods when employed on the dataset.

Evaluation is based on the success rate of the segments, which shows the number of charts whose all textual and graphical segments were extracted without failure with respect to the total number of charts [14]. The results are treated as successful only if the system predicts all segments without missing any segment.

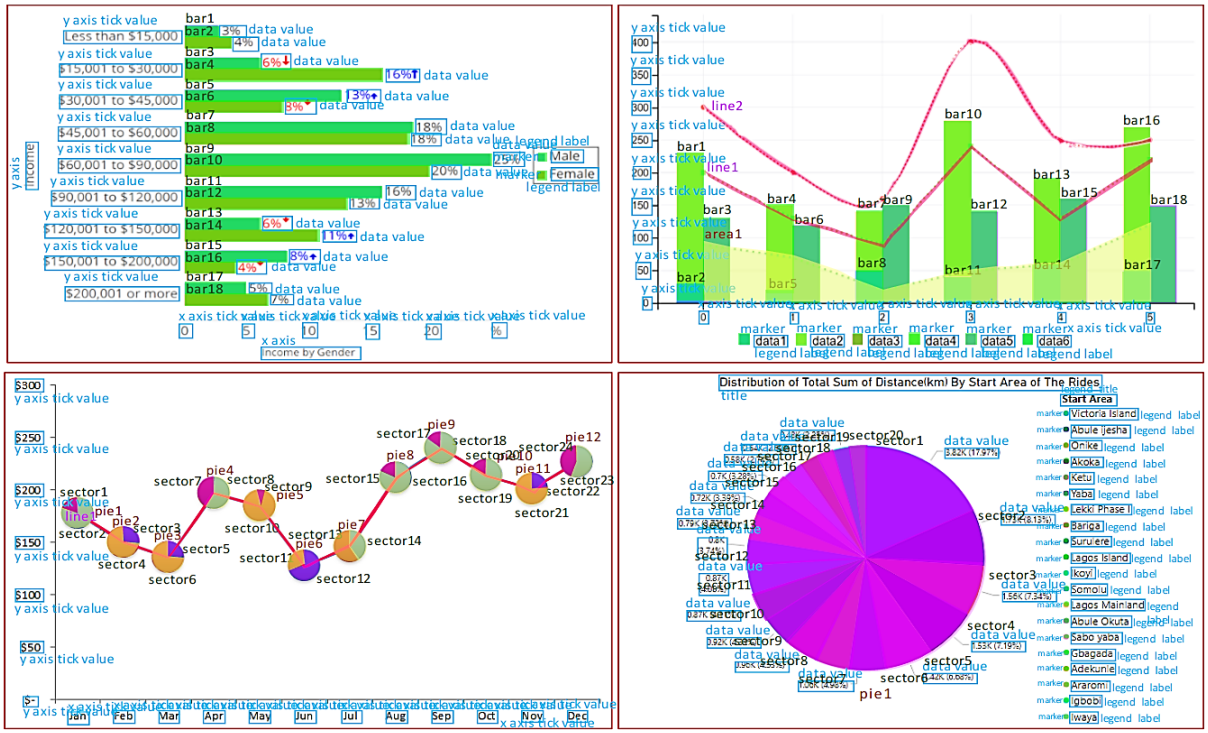


Figure 9: A few samples of detection and role classification

Table 6: Performance evaluation (in %) for data extraction with and without relationship graph

Chart Types	With Relation Building	Without Relation Building
Vertical Bar Chart	98.12	86.68
Line Chart	96.63	82.82
Group Bar Charts	96.98	84.69
Histogram	98.73	79.22
Pie Chart	96.21	85.25
Area Chart	93.50	80.35
Stack Area Chart	95.72	83.21
Horizontal Bar Chart	98.50	86.96

Table 7: Accuracy rate (%) for graphical components

Types ↓	Proposed Model				Savva et al. [11]				Choi et al. [10]				Jung et al. [17]				Poco & Heer [8]			
Parameters*	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Line Chart	100	100	98	96	92	92	90	92	96	98	94	95	93	94	96	96	95	96	92	93
Vertical Bar Chart	100	100	96	98	90	90	92	89	97	96	95	96	98	95	96	98	97	95	95	96
Horizontal Bar Chart	100	100	96	96	85	80	85	82	96	98	94	93	91	93	94	96	95	96	93	93
Histogram	100	100	94	98	90	90	90	91	98	97	92	96	92	93	93	98	98	97	91	94
Pie Chart	94	-	96	94	85	-	95	92	93	-	96	92	91	-	90	90	-	-	-	-
Group Bar Chart	94	96	92	92	92	92	89	86	92	94	90	91	92	93	90	91	90	91	88	85
Area Chart	100	98	92	90	93	93	85	86	-	-	-	-	93	93	92	88	97	94	90	91
Stack Area Chart	100	98	88	92	91	90	89	82	-	-	-	-	96	93	93	89	98	94	85	82

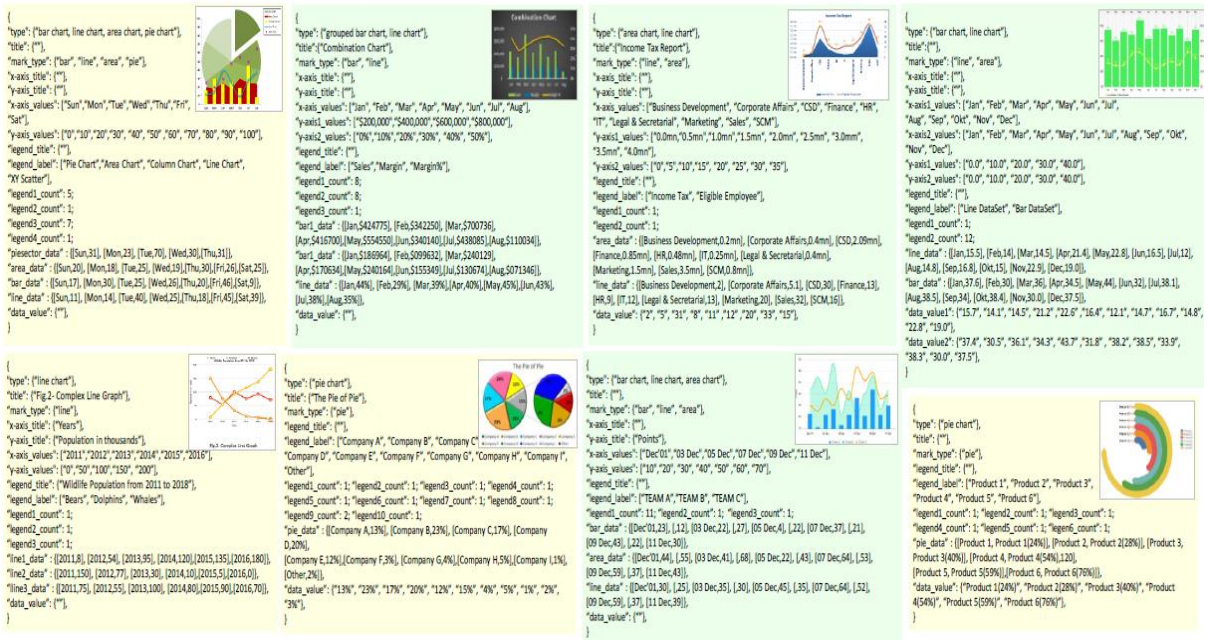


Figure 10: A few samples of extracted data encoding and data interpretation

Summary: The chapter proposes a semi-automated system for chart data extraction and interpretation, focusing on textual and graphical components individually and building the relationships between every chart component. Moreover, the proposed approach also identifies, interprets, and maps legend values with the corresponding graphical components, along with scale mapping and axis data types inferencing, unlike benchmark methods discussed in the literature, which tend to avoid legend analysis.

Heuristic and hand-coded methods were employed to extract direct and implied data from chart images. Once these visualization data are made accessible to the machine, it can be used in multiple visualization applications such as redesigning charts, summarization, data analytics, retrieval, and question-answering system. On the other hand, the underlying data should contain relevant knowledge and accurate information to aid the machine in making an informed decision. The details obtained from the proposed approach are further used for generating concise summaries, which is discussed in Chapter 6.

Chapter 6: Chart Summarization

6.1. Background

Each chart conveys an intended message of the graphic's high-level content and portrays the abstract meaning of the data graphics. Intended messages are also stated as "*how the graphics should be understood"* by the viewer [21]. Different charts effectively illustrate meanings or factual data using vibrant and diverse graphical components. Summarization is generally stated as the gist of message content or subject matter. Various eminent screen reading software is available in the market, efficiently proffering textual components from documents. Moreover, it fails to deliver the required contents from the graphical components. These screen readers skip charts and such related images. An assistive reader of the document needs to get the high-level information imparted by the chart images. Moreover, the graphics contribute to document understanding and hence cannot be ignored. Thus, extracting hidden meaning from chart images becomes necessary to deduce the global message on underlying data depicted by charts [17].

6.2. Pilot Study

The primary objective is to provide an assistive aid to visually impaired persons with a system that interprets and summarizes scientific charts. Hence, before designing a summary generation system, it is necessary to interact with the people with disability about their needs and requirement for generated summaries. A few rehabilitation centers, NGOs, and ophthalmologists who offer a helping hand to people with disabilities were contacted. While interacting with these organizations and ophthalmologists, the issues faced by the visually impaired while learning from documents and e-books were identified.

Participants mostly encounter data table tables or scientific images such as bar charts, line charts, pie charts, histograms, and sometimes area charts are mostly used in subject books. Participants and doctors feel there are no mechanisms to read and understand scientific images while reading a document, book, or website. All participants, especially school students, agreed that the screen readers should support image description as it will benefit them while reading coursebooks. The observations from the pilot study inclined towards the design of the proposed summary generation system.

6.3. Proposed Method: Automated Summary Generation using Deep Learning Models[5]

A teacher-student distillation learning approach is proposed for summary generation, where the teacher model guides the student. The student model extracts linguistic knowledge to determine the essential relations and sentences from the teacher output summary, generating detailed and short summaries from the input chart image, following the supervised learning approach. The proposed summary generation model (shown in Fig. 11) comprises semantic embedding between chart elements based on their positional encodings, the relation network responsible for determining the relationships between data and chart elements, and the LSTM model that generates the final summary.

Given image features and positional encoding of elements, semantic embedding finds the sets of textual elements that are most relevant to corresponding graphical elements. The ((x, y), (cx, cy)) coordinate of each graphical segment is fused with semantic embedding and relation network to determine the hidden or abstract meaning of the chart, where (x, y) is the bounding box values and (cx, cy) is the chart coordinates of the graphical component. Interpretation of intended summary is based upon the graphical segment inter-relations and graphical-textual segment semantic relation. Hence, the graphical (object) pair is conditioned on its semantic embedding with textual elements. Pair-wise relation mechanism aids in determining the relation between graphical elements. The role of the teacher model is to generate a detailed summary, while the student model generates both detailed and short

[5]P Mishra, S Kumar, M K Chaube, "A Style-Invariant Intended Summarization for Charts", IEEE Transactions of Multimedia.

summaries. Short summaries portray only the intended meaning and crisp details of the image. The teacher model produces a soft distribution of the output summary, elevating and refining the student output summary based on the base summary.

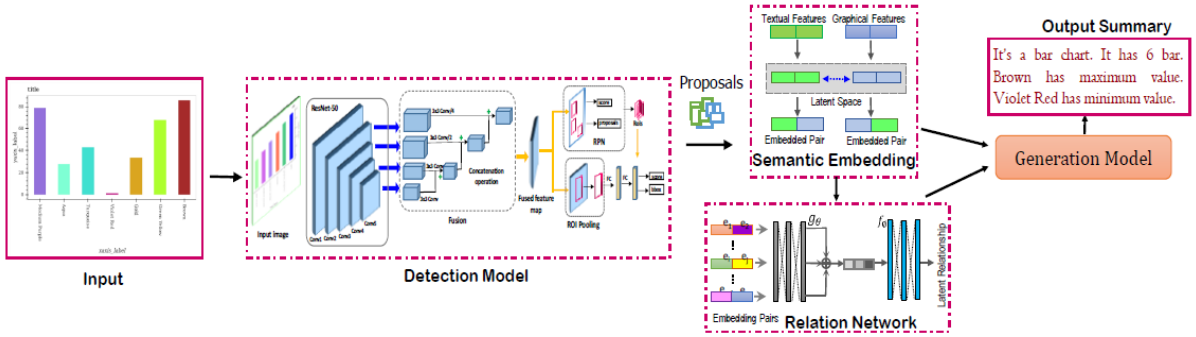


Figure 11: Automated Summary Generation Pipeline

6.4. Experimental Results and Analysis

This section portrays the task-based and metric-based evaluation of the generated summaries.

6.4.1. Datasets Used

- VegaLite [31]
- Hand-drawn Chart Sketch
- Chart Sketches

6.4.2. Results Obtained

For Table 8, the performance of the proposed model is compared with benchmark summary generation models, where for better evaluation of the impact of each module over summaries, the experiments were separately performed employing the following model over considered datasets,

- DM + SM: Detection model as encoder and LSTM as decoder.
- DM + SE + SM: Detection model as encoder, LSTM as decoder, semantic embedding between elements.
- DM + RN + SE + SM: Detection model as encoder, LSTM as decoder, semantic embedding and relation network between elements.

where, DM stands for component detection model, SM stands for adaptive summarization model, SE is semantic embedding, and RN is relational network. The results show that the student model performs significantly better than other models, even surpassing the teacher model. Fig. 12 shows the summaries generated for sample charts.

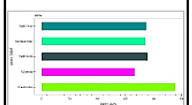
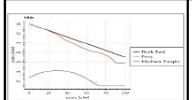
Input Image	Teacher Summary	Student Summary	
		Detailed	Short
 FigureQA	It is bar chart having 5 horizontal bar. The chart has xaxis_label on x axis and yaxis_label on y axis. X axis lies between 0 to 100 and y axis lies between Chartreuse to Dark Cyan. Chartreuse is the longest bar having xaxis_label value 96. Magenta is the smallest bar having xaxis_label value 66. Dark State is the second largest bar having xaxis_label value 76. Medium Mint and Dark Cyan has equal xaxis_label value 74.	It is horizontal bar chart having title title. The chart has xaxis_label on x axis ranging from 0 to 100 and yaxis_label on y axis ranging from Chartreuse to Dark Cyan. It has 5 bars. Chartreuse has the highest xaxis_label 96 while Magenta has lowest xaxis_label 66. Medium Mint and Dark Cyan has same xaxis_label 74. All bars except Chartreuse show steady trend.	Chartreuse is the longest bar and Magenta is the smallest bar. Dark State is the second largest bar. Medium Mint and Dark Cyan are equal but smaller than Dark State. yaxis_label shows fall in trend and with slight increase lies steady.
 FigureQA	It is line chart having 3 lines. The chart has xaxis_label on x axis and yaxis_label on y axis. X axis lies between 0 to 100 and y axis lies between 82 to 94. It has legend as Dark Red, Peru and Medium Purple. Dark Purple falls continuously from yaxis_label 92.5 to 87 at xaxis_label 100. Peru drastically falls from yaxis_label 94 to 88.5 at xaxis_label 60. Peru gradually reduces up to yaxis_label 87.5 at xaxis_label 90 and lies steady further. Medium Purple rises from yaxis_label 83 to 84.5 at xaxis_label 30 then falls to yaxis_label 80.5 at xaxis_label 50 and lies steady further.	It is line chart having title title. The chart has xaxis_label on x axis ranging from 0 to 100 and yaxis_label on y axis ranging from 82 to 94. It has 3 lines. It has 3 legend as Dark Red, Peru and Medium Purple. Dark Purple has yaxis_label 92.5 at xaxis_label 0 then continuously falls up to 87 at xaxis_label 100. Peru has highest yaxis_label 94 at xaxis_label 0 which drastically falls till yaxis_label 87.5 at xaxis_label 90 and lies steady till xaxis_label 100. Medium Purple rises from yaxis_label 83 to 84.5 at xaxis_label 30 which falls to yaxis_label 80.5 at xaxis_label 50 and lies steady till xaxis_label 100.	Dark Purple shows steep fall. Peru has highest yaxis_value. Peru gradually drops then rises again slightly and then drops again. Medium Purple shows a rise which continuously falls then remains steady.

Figure 12: Summaries generated from teacher and student model

Table 8: Evaluation results of generated summaries using metrics over different datasets

Dataset	Model	Evaluation Metrics						
		CIDEr	BLEU1	BLEU2	BLEU3	BLEU4	METEOR	ROUGE
FigureQA	DM+SM	0.62	0.38	0.29	0.25	0.21	0.22	0.44
	DM+SE+SM	0.79	0.42	0.33	0.29	0.24	0.28	0.49
	DM+SE+RN+SM	0.90	0.51	0.39	0.38	0.29	0.37	0.88
	Chen et al.[22]	0.81	0.46	0.31	0.35	0.24	0.29	0.53
	VD-SAN [23]	0.69	0.41	0.29	0.37	0.26	0.25	0.48
	Qian et al. [32]	0.56	0.34	0.18	0.21	0.12	0.20	0.35
VegaLite	DM+ST	0.72	0.42	0.41	0.35	0.29	0.37	0.67
	DM+SE+SM	0.81	0.56	0.49	0.42	0.34	0.42	0.79
	DM+SE+RN+SM	0.91	0.67	0.55	0.53	0.41	0.49	0.91
	Chen et al.[22]	0.81	0.61	0.49	0.48	0.38	0.45	0.76
	VD-SAN [23]	0.79	0.57	0.42	0.41	0.29	0.37	0.61
	Qian et al. [32]	0.66	0.46	0.33	0.39	0.22	0.28	0.48
Web Chart	DM+SM	0.54	0.37	0.20	0.30	0.32	0.20	0.42
	DM+SE+SM	0.64	0.46	0.39	0.49	0.37	0.34	0.72
	DM+SE+RN+SM	0.89	0.62	0.46	0.51	0.51	0.55	0.88
	Chen et al. [22]	0.72	0.48	0.40	0.45	0.42	0.49	0.75
	VD-SAN [23]	0.66	0.37	0.31	0.42	0.36	0.33	0.66
	Qian et al. [32]	0.48	0.40	0.37	0.31	0.35	0.33	0.60
Hand drawn	DM+SM	0.58	0.41	0.24	0.33	0.35	0.31	0.49
	DM+SE+SM	0.77	0.59	0.31	0.39	0.42	0.47	0.69
	DM+SE+RN+SM	0.95	0.68	0.52	0.57	0.57	0.61	0.94
	Chen et al.[22]	0.61	0.50	0.40	0.35	0.46	0.59	0.72
	VD-SAN [23]	0.49	0.40	0.31	0.30	0.38	0.43	0.55
	Qian et al. [32]	0.32	0.20	0.28	0.20	0.23	0.38	0.43
Chart Sketch	DM+SM	0.61	0.42	0.30	0.33	0.33	0.28	0.43
	DM+SE+SM	0.78	0.55	0.41	0.51	0.36	0.46	0.75
	DM+SE+RN+SM	0.91	0.61	0.47	0.56	0.46	0.57	0.89
	Chen et al.[22]	0.77	0.53	0.41	0.47	0.40	0.41	0.74
	VD-SAN [23]	0.70	0.48	0.36	0.40	0.31	0.39	0.61
	Qian et al. [32]	0.35	0.42	0.37	0.39	0.30	0.33	0.58

A task-based evaluation is also performed with people with visual impairment to evaluate the usefulness of the generated summaries, i.e., are users able to answer important questions regarding the knowledge conveyed by the summaries and compare their performance to that of sighted users viewing the same summaries. Mentors/teachers from each organization were asked to select the chart images on which they wanted the proposed summary generation system to generate summaries. They were made aware of the pilot study and chart types in consideration. On average, 228 questions were answered by the participants overall, out of which 205 questions were correctly answered.

A feedback session was conducted with the group members when the experiment concluded. Participants said that summaries cover all the required and correct details of the charts, and details are presented in well-sorted order making summaries easily understandable. Few members gave some suggestions for the system improvements, such as members suggested i) the mobile-based system should also be made available so that participants using screen readers on smartphones can get access to scientific images, ii) the system should include summarizing data tables, flow charts, and figures, iii) the system should be extended for people with deaf-blind impairment using Braille devices.

Summary: The summary generation system produces concise summaries from extracted data for different types of charts. The key challenge was to extract correct direct and implied data from the charts to generate semantically correct summaries. The proposed approaches generated an intended message from chart images using direct and derived data from charts and their relationship. The proposed approach focuses on chart types with varied background colors, styles, and appearances. The automated system successfully covers all details of charts, followed by the intention message stating “what message is conveyed in the chart and how it should be understood.”

Chapter 7: Conclusion and Future Scope

study explores the chart processing pipeline as a modular system and attempts to overcome the underlying challenges of processing diverse, complex chart images. The underlying data, when extracted correctly, could be used in multiple areas of data analysis and visualization. A comprehensive literature review reveals the variety of work in text data extraction from charts and depicts the open challenges in automated data identification and interpretation, such as arbitrary text location, variable size graphics, and variations in axis orientations. There has been extensive work that focuses on chart classification targeting different types of charts. There is a strong need to address the challenging issues of chart processing by stepping away from heuristics, hard assumptions, and constraints while progressing towards robust, style-invariant and scalable methods that can handle diverse chart types all under one roof. This study has formulated four distinct problem modules related to chart detection, similar structure classification, data extraction and summary generation. Six style-independent supervised approaches are proposed to solve these problems that target different chart processing pipeline issues. Consequently, the conceptual and experimental study yielded some important observations which are described in the chapters. The work also proposed three chart datasets MixedBad, Handcrafted chart images, and chart sketches.

Although the proposed approaches have succeeded in overcoming the challenges of chart processing and limitations of existing works, there is a scope for improvement in each module. For automated data extraction and interpretation, the learning approach needs to address the diversity in style and structure of charts, such as info-graphics charts which are highly abstract with impactful visualization patterns. Processing and interpreting dynamic charts, combined charts and complex charts (having comparison variables of more than 20) are still open issues that require imperative future efforts. Thus, attending to the looming interest in the expanses of chart processing and interpretation, a few of the open challenges such as data extraction from charts having patterns or images as background, inclined chart interpretation, processing dimensions, and interpreting data correctly from 3D charts and combined charts are taken as the future scope of the work. Also, the summary generation system will be extended over other chart types, such as box plots, scatter charts, and concentric pie charts, with additional support for other local Indian languages.

Organization of Thesis

A detailed description of each chapter is given below:

Chapter-1
This chapter discusses various types of charts, a brief history of charts, the background of the chart processing system, various challenges of chart processing and summarization, and various imminent applications, motivations, and objectives of the work.

Chapter-2
This chapter discusses existing literature and state-of-the-art methods for chart recognition, classification, data extraction, summarization, and limitations in existing works. It also discusses chart processing tools, existing datasets, and evaluation mechanisms.

Chapter-3
This chapter focuses on various document layout analyses and detecting chart images from multi-lingual document images. The chapter proposes a cross-domain document object detection model which trains on labeled document images and adapts to unlabelled multi-lingual heterogeneous document images. The proposed model is compared with state-of-the-art detection models employed over combinations of source and target document image datasets.

Chapter-4
This chapter explores the visualization attributes of the charts, preliminary testing for classification models over different chart types, and discusses within-class and between-class chart classification challenges. The chapter proposes fusion-based chart classification methods using similar and dissimilar microstructural features. The proposed method is compared with existing benchmark methods and models over different types and sub-types of chart datasets.

Chapter-5
This chapter discusses the pipeline of chart data extraction and interpretation, the structural challenges of data extraction, and the importance of the relationship between chart components. The chapter proposes a semi-automated pipeline for extracting direct and derived data from charts, extracting graphical components, and interpreting the extracted data using inter-component relationships of components. Compared with state-of-the-art literature methods over different chart datasets, the proposed model surpasses existing benchmark methods.

Chapter-6
This chapter discusses the current summary types of literature, prerequisites of the summaries, and a pilot study to understand the need for the summary generation system. The chapter proposes a summary generation system that fuses semantic embeddings, positional encodings, and inter-component relationships with the domain-adaptive model to generate short summaries in English and Hindi languages. The summaries are evaluated using task-based evaluation mechanisms and various other metrics compared with a few existing summarization models.

Chapter-7
This chapter describes the conclusion drawn from the entire work, mentioning the limitations of the work and the future scope within the chart processing research domain.

List of Publications

Published Articles

1. **P Mishra**, S Kumar, M K Chaube, "Graph Interpretation, Summarization and Visualization Techniques: A Review and Open Research Issues", Special Issue of Springer Multimedia Tools and Applications, 2022. https://doi.org/10.1007/s11042-021-11582-9.
2. **P Mishra**, S Kumar, M K Chaube, "ChartFuse: a novel fusion method for chart classification using heterogeneous microstructures", Springer Multimedia Tools and Applications, 80, no. 7, 2021. https://doi.org/10.1007/s11042-020-10186-z.
3. **P Mishra**, S Kumar, M K Chaube, "Classifying Chart Based on Structural Dissimilarities using Improved Regularized Loss Function", Springer Neural Processing Letters, 2021. https://doi.org/10.1007/s11063-021-10735-z.
4. **P Mishra**, S Kumar, M K Chaube, "Dissimilarity-Based Regularized Learning of Charts", ACM Transactions on Multimedia Computing, Communications, and Applications, 17, no. 4, 2021. https://doi.org/10.1145/3458884.
5. **P Mishra**, S Kumar, M K Chaube, "Evaginating scientific charts: Recovering direct and derived information encodings from chart images", Springer Journal of Visualization, 2021. https://doi.org/10.1007/s12650-021-00800-z.
6. **P Mishra**, S Kumar, M K Chaube, U Shrawankar, "ChartVi: Charts Summarizer for Visually Impaired", Elsevier Journal of Computer Languages, Volume 69, April 2022. https://doi.org/10.1016/j.cola.2022.101107.
7. **P Mishra**, S Kumar, M K Chaube, "Adaptive Learning for Cross-domain Document Illustrative Detection Using Dual Attention Mechanism", Springer Multimedia Information Retrieval (Revision Submitted).

Communicated Articles

8. **P Mishra**, S Kumar, M K Chaube, "A Style-Invariant Intended Summarization for Charts", IEEE Transactions of Multimedia (In review).

Published Conferences

1. **P Mishra**, S Kumar, M K Chaube, "Assistive Accessible Charts for Visually Impaired Students: An Automated Learning System", 36th Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence, Canada, Feb 28-March 2 2022.
2. **P Mishra**, S Kumar, M K Chaube, "Dissimilarity Based Regularized Deep Learning Model for Information Charts", Published in IEEE 4th International Conference on Imaging, Vision & Pattern Recognition (IVPR), Japan, 2020. pp. 1-6.
3. **P Mishra**, S Kumar, M K Chaube, "Interpretation and Segmentation of Chart Images Using h-Means Image Clustering Algorithm", Springer 4th International Conference on Data Management, Analytics and Innovation (ICDMAI), 2019, vol 1174. Springer, Singapore.

References

[1] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," IEEE transactions on visualization and computer graphics, vol. 16, no. 6, pp. 1139–1148, 2010.

[2] Li Kai, C. Wigington, C. Tensmeyer, H. Zhao, N. Barmpalios, V. I. Morariu, V. Manjunatha, T. Sun, and Y. Fu, "Cross-domain document object detection: Benchmark suite and method," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12 915–12 924.

[3] P. Lee, J. D. West, and B. Howe, "Viziometrics: Analyzing visual information in the scientific literature," IEEE Transactions on Big Data, vol. 4, no. 1, pp. 117–129, 2017.

[4] W. Huang and C.-L. Tan, "Locating charts from scanned document pages," in Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), vol. 1. IEEE, 2007, pp. 307–311.

[5] Z. Chen, M. Cafarella, and E. Adar, "Diagramflyer: A search engine for data driven diagrams," in Proceedings of the 24th International conference on world wide web, 2015, pp. 183–186.

[6] S. Demir, S. Carberry, and K. F. McCoy, "Summarizing information graphics textually," Computational Linguistics, vol. 38, no. 3, pp. 527–574, 2012.

[7] C. Chen, R. Zhang, E. Koh, S. Kim, S. Cohen, and R. Rossi, "Figure captioning with relation maps for reasoning," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 1537–1545.

[8] J. Poco and J. Heer, "Reverse-engineering visualizations: Recovering visual encodings from chart images," in Computer Graphics Forum, vol. 36, no. 3. Wiley Online Library, 2017, pp. 353–363.

[9] S. E. Kahou, V. Michalski, A. Atkinson, Á. Kádár, A. Trischler, and Y. Bengio, "Figureqa: An annotated figure dataset for visual reasoning," arXiv preprint arXiv:1710.07300, 2017.

[10] J. Choi, S. Jung, D. G. Park, J. Choo, and N. Elmqvist, "Visualizing for the non-visual: Enabling the visually impaired to use visualization," in Computer Graphics Forum, vol. 38, no. 3. Wiley Online Library, 2019, pp. 249–260.

[11] M. Savva, N. Kong, A. Chhajta, L. Fei-Fei, M. Agrawala, and J. Heer, "Revision: Automated classification, analysis and redesign of chart images," in Proceedings of the 24th annual ACM symposium on User interface software and technology, 2011, pp. 393–402.

[12] Siddiqui, Shoaib Ahmed, Muhammad Imran Malik, Stefan Agne, Andreas Dengel, and Sheraz Ahmed. "Decnt: Deep deformable cnn for table detection." IEEE Access 6 (2018): 74151-74161.

[13] Singh, Pranaydeep, Srikrishna Varadarajan, Ankit Narayan Singh, and Muktabh Mayank Srivastava. "Multi-domain Document Layout Understanding Using Few-Shot Object Detection." In International Conference on Image Analysis and Recognition, pp. 89-99. Springer, Cham, 2020.

[14] Agarwal, Madhav, Ajoy Mondal, and C. V. Jawahar. "Cdec-net: Composite deformable cascade network for table detection in document images." In 2020 25th International Conference on Pattern Recognition (ICPR), pp. 9491-9498. IEEE, 2021.

[15] N. Vo, K. Nguyen, T. Nguyen, and K Nguyen. Ensemble of deep object detectors for page object detection. In Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication, IMCOM '18, pages 11:1– 11:6, New York, NY, USA, 2018. ACM.

[16] Sarshogh, Mohammad Reza, and Keegan Hines, A Multi-task Network for Localization and Recognition of Text in Images, IEEE International Conference on Document Analysis and Recognition, (2019) 494-501. 10.1109/gao.2019.00085.

[17] Jung, Daekyoung, Wonjae Kim, Hyunjoo Song, Jeong-in Hwang, Bongshin Lee, Bohyoung Kim, and Jinwook Seo, ChartSense: Interactive data extraction from chart images, ACM CHI Conference on Human Factors in Computing Systems, (2017) 6706-6717.

[18] Nair, Rathin Radhakrishnan, Nishant Sankaran, Ifeoma Nwogu, and Venu Govindaraju, Understanding line plots using Bayesian Network, 12th IAPR Workshop on Document Analysis Systems (DAS), (2016) 108-113. /10.1109/DAS.2016.73.

[19] Cliche, Mathieu, et al., Scatteract: Automated extraction of data from scatter plots, Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 10534 (2017).

[20] Mahmood, Aqsa, Imran Sarwar Bajwa, and Kiran Qazi, An automated approach for interpretation of statistical graphics, Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, vol. 2, pp. 376-379. IEEE, 2014.

[21] Demir, Seniz, David Oliver, Edward Schwartz, Stephanie Elzer, Sandra Carberry, Kathleen F. Mccoy, and Daniel Chester, Interactive SIGHT: textual access to simple bar charts., New Review of Hypermedia and Multimedia 16 (2010) 245-279.

[22] Chen, C.; Zhang, R.; Koh, E.; Kim, S.; Cohen, S.; and Rossi, R. 2020. Figure Captioning with Relation Maps for Reasoning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 1537–1545.

[23] X. He, Y. Yang, B. Shi, and X. Bai, "VD-SAN: Visual-densely semantic attention network for image caption generation," Neurocomputing, vol. 328, pp. 48–55, 2019.

[24] M. Li, Y. Xu, L. Cui, S. Huang, F.Wei, Z. Li, and M. Zhou, "Docbank: A benchmark dataset for document layout analysis," arXiv preprint arXiv:2006.01038, 2020.

[25] X. Zhong, J. Tang, and A. J. Yepes, “Publaynet: largest dataset ever for document layout analysis,” in 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019, pp. 1015–1022.
[26] K. Davila, B. U. Kota, S. Setlur, V. Govindaraju, C. Tensmeyer, S. Shekhar, and R. Chaudhry, “Icdar 2019 competition on harvesting raw tables from infographics (chart-infographics),” in 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019, pp. 1594–1599.
[27] A. Mondal, P. Lipps, and C. Jawahar, “Iiit-ar-13k: a new dataset for graphical object detection in documents,” in International Workshop on Document Analysis Systems. Springer, 2020, pp. 216–230.
[28] L. Gao, X. Yi, Z. Jiang, L. Hao, and Z. Tang, “Icdar2017 competition on page object detection,” in 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1. IEEE, 2017, pp. 1417–1422.
[29] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” International conference on machine learning. PMLR, 2017, pp. 1126–1135.
[30] K. Li, C. Wigington, C. Tensmeyer, V. I. Morariu, H. Zhao, V. Manjunatha, N. Barmpalios, and Y. Fu, “Rpcl: A framework for improving cross-domain detection with auxiliary tasks,” arXiv preprint arXiv:2104.08689, 2021.
[31] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, “Vega-lite: A grammar of interactive graphics,” IEEE transactions on visualization and computer graphics, vol. 23, no. 1, pp. 341–350, 2016.
[32] X. Qian, E. Koh, F. Du, S. Kim, J. Chan, R. A. Rossi, S. Malik, and T. Y. Lee, “Generating accurate caption units for figure captioning,” in Proceedings of the Web Conference 2021, 2021, pp. 2792–2804.
[33] X. Yang, E. Yumer, P. Asente, M. Kraley, D. Kifer, and C. Lee Giles, “Learning to extract semantic structure from documents using multimodal fully convolutional neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5315–5324.